PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-055693
(43)Date of publication of application : 20.02.2002

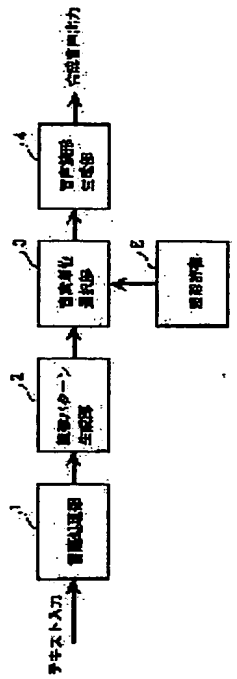
(51)Int.Cl. G10L 13/06

(21)Application number : 2000-242068
(22)Date of filing : 10.08.2000

(71)Applicant : SANYO ELECTRIC CO LTD
(72)Inventor : HIRAI HIROYUKI

(54) METHOD FOR SYNTHESIZING VOICE

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a method for synthesizing a voice with which a phonemic piece with an inferior quality causing sound quality degradation is hardly selected as an optimum phonemic piece, without having to significantly correct waveform dictionary.
SOLUTION: The method for synthesizing voice consists of a step, in which penalty information is added to supplementary information for each phonemic unit, a step in which a user input a synthesized voice part of inferior quality when the user decides the quality of a synthesized voice as being inferior as a result of hearing, and a step in which the value for enlarging a calculated value distorted from a target forcibly, when the phonemic piece corresponding to the synthesized voice part of inferior quality is selected as a candidate is set to penalty information in the supplementary information of the above phonemic piece, when the synthesized voice part of inferior quality inputted by the user is inputted.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2002－55693
（P2002－55693A）
(43)公開日　平成14年２月20日(2002.2.20)

(51)Int.Cl.⁷	識別記号	ＦＩ	ﾃｰﾏｺｰﾄﾞ(参考)
Ｇ１０Ｌ　13/06		Ｇ１０Ｌ　5/04	Ｆ　５Ｄ０４５

審査請求　有　　請求項の数２　ＯＬ　（全６頁）

(21)出願番号　特願2000－242068(P2000－242068)

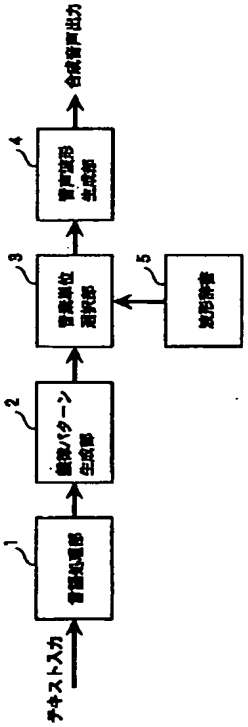
(22)出願日　平成12年８月10日(2000.8.10)

(71)出願人　000001889
三洋電機株式会社
大阪府守口市京阪本通２丁目５番５号
(72)発明者　平井　啓之
大阪府守口市京阪本通２丁目５番５号　三洋電機株式会社内
(74)代理人　100086391
弁理士　香山　秀幸

Ｆターム(参考)　5D045 AA20

(54)【発明の名称】　音声合成方法

(57)【要約】
【課題】　この発明は、波形辞書の大幅な修正を行なうことなく、音質劣化につながる品質の悪い音素片が最適な音素片として選択されにくくすることができる音声合成方法を提供することを目的とする。
【解決手段】　各音素単位の補助情報にペナルティ情報を追加しておくステップ、ユーザが音声合成結果を聞いて、その品質が悪い場合には、品質の悪い合成音声箇所をユーザに入力させるステップ、ならびにユーザによって入力された品質の悪い合成音声箇所が入力された場合には、当該品質の悪い合成音声箇所に対応する音素片の補助情報内のペナルティ情報に、当該音素片が候補として選択されたときにターゲットとの歪み算出値を強制的に大きくさせるような値を設定するステップを備えている。



1

【特許請求の範囲】

【請求項１】　複数の音声単位と各音素単位毎にターゲットとの歪みを算出するために用いられる補助情報とが波形辞書に格納されており、波形辞書に格納されている音素単位の組み合わせの中で、ターゲットとの歪みが最も少なくなる組み合わせを選択する音素単位選択型の音声合成方法において、

各音素単位の補助情報にペナルティ情報を追加しておくステップ、

ユーザが音声合成結果を聞いて、その品質が悪い場合には、品質の悪い合成音声箇所をユーザに入力させるステップ、ならびにユーザによって入力された品質の悪い合成音声箇所が入力された場合には、当該品質の悪い合成音声箇所に対応する音素片の補助情報内のペナルティ情報に、当該音素片が候補として選択されたときにターゲットとの歪み算出値を強制的に大きくさせるような値を設定するステップ、

を備えていることを特徴とする音声合成方法。

【請求項２】　複数の音声単位と各音素単位毎にターゲットに対する適応度を算出するために用いられる補助情報とが波形辞書に格納されており、波形辞書に格納されている音素単位の組み合わせの中で、ターゲットに対する適応度が最も大きくなる組み合わせを選択する音素単位選択型の音声合成方法において、

各音素単位の補助情報に優先度情報を追加しておくステップ、

ユーザが音声合成結果を聞いて、その品質が悪い場合には、品質の悪い合成音声箇所をユーザに入力させるステップ、ならびにユーザによって入力された品質の悪い合成音声箇所が入力された場合には、当該品質の悪い合成音声箇所に対応する音素片の補助情報内の優先度情報に、当該音素片が候補として選択されたときにターゲットに対する適応度の算出値を強制的に小さくさせるような値を設定するステップ、

を備えていることを特徴とする音声合成方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】この発明は、任意のテキスト情報を合成音声で読み上げることのできる音声合成方法に関する。

【０００２】

【従来の技術】図１は、音声合成装置の概略構成を示している。

【０００３】入力された日本語仮名漢字混じりのテキストは、言語処理部１で形態素解析、係り受け解析が行なわれ、音素記号、アクセント記号等に変換せしめられる。

【０００４】韻律パターン生成部２では、音素記号、アクセント記号列および形態素解析結果から得られる入力テキストの品詞情報を用いて、音韻継続時間長（声の長さ DUR^{T} ）、基本周波数（声の高さ $F0^{T}$ ）、母音中心のパワー（声の大きさ POW^{T} ）等の推定が行なわれる。

【０００５】音素単位選択部３では、推定された音韻継続時間長 DUR^{T} 、基本周波数 $F0^{T}$ および母音中心のパワー POW^{T} に最も近く、かつ波形辞書５に蓄積されている音素単位（音素片）を接続したときの歪みが最も小さくなる音素片の組み合わせがＤＰ（動的プログラミング）を用いて選択される。

【０００６】音声波形生成部４では、選択された音素片の組み合わせにしたがって、ピッチを変換しつつ音素片の接続を行なうことによって音声が生成される。

【０００７】図２は、波形辞書５の内容を示している。波形辞書５は、複数の音素片が格納された音素片格納部５１と、音素片格納部５１内の各音素片に関する補助情報が格納された補助情報格納部５２とがある。補助情報には、音素片のパワー（POW^{Dic} ）、基本周波数（ $F0^{Dic}$ ）、継続時間長（ DUR^{Dic} ）等がある。

【０００８】ところで、音素単位選択部３では、波形辞書５に蓄積されている音素片の組み合わせの中で、歪みが少なくなる組み合わせを選択しているが、この歪みには次のようなものがある。

【０００９】つまり、図３に示すように、u_{i-1} 、u_i 、u_{i+1} を波形辞書５から抽出した音素片として、t_{i-1} 、t_i 、t_{i+1} を実際に使用する環境（ターゲット）とすると、ui に対する歪みには、C_i^{t} と、C_i^{c} とがある。

【００１０】ここで、C_i^{t} は、ｉ番目の音素について辞書から抽出した音素片（u_i ）と実際に使用する環境（ターゲット t_i ）との間の歪みである。また、C_i^{c} は、ｉ番目の音素片（u_i ）と、ｉ－１番目の素片（u_{i-1} ）とを接続したときに生じる歪みである。音素単位選択部３は、動的計画法（ＤＰ法）に用いて音素片を接続していき、入力された全ての音素に対する C_i^{t} と C_i^{c} との総和 C^{all} が最小となる素片の組み合わせを選択する。

【００１１】C_i^{t} は、次の数式１で表される。

【００１２】

【数１】

$$C_i^{t} = w'_{POW} D'_{POW}(t_i, u_i) + w'_{F0} D'_{F0}(t_i, u_i) + w'_{DUR} D'_{DUR}(t_i, u_i)$$

【００１３】数式１において、各変数は、次のように定義される。

【００１４】D_{POW}^{t} （t_i , u_i ）は、ｉ番目の音素について、辞書から抽出した音素片（u_i ）のパワー（POW^{Dic} （i））と、実際に使用する環境（ターゲット t_i ）のパワー（POW^{T} （i））との間の距離の自乗であり、｛（POW^{Dic} （i））－（POW^{T} （i））｝^{2} となる。

【００１５】w_{POW}^{t} は、D_{POW}^{t} （t_i , u_i ）に対する重み係数である。

【００１６】D_{F0}^{t} （t_i , u_i ）は、ｉ番目の音素について、辞書から抽出した音素片（u_i ）の基本周波数

（FO^{Dic}（i））と、実際に使用する環境（ターゲットt_i）の基本周波数（FO^{T}（i））との間の距離の自乗であり、｛（FO^{Dic}（i））－（FO^{T}（i））｝^{2}となる。

【００１７】w_{FO}^{c}は、D_{FO}^{c}（t_i，u_i）に対する重み係数である。

【００１８】D_{DUR}^{c}（t_i，u_i）は、i番目の音素について、辞書から抽出した音素片（u_i）の継続時間長（DUR^{Dic}（i））と、実際に使用する環境（ターゲットt_i）の継続時間長（DUR^{T}（i））との間の距離の自乗であり、｛（DUR^{Dic}（i））－（DUR^{T}（i））｝^{2}となる。

【００１９】w_{DUR}^{c}は、D_{DUR}^{c}（t_i，u_i）に対する重み係数である。

【００２０】C_i^cは、次の数式２で表される。

【００２１】

【数２】

$$C_i^c = w_{POW}^c D_{POW}^c(u_i, u_{i-1}) + w_{FO}^c D_{FO}^c(u_i, u_{i-1}) + w_{SPC}^c D_{SPC}^c(u_i, u_{i-1})$$

【００２２】数式２において、各変数は、次のように定義される。

【００２３】D_{POW}^{c}（u_i，u_{i-1}）は、i番目の音素片（u_i）の始端のパワー（POW^{DicS}（i））と、i－１番目の音素片（u_{i-1}）の終端のパワー（POW^{DicE}（i-1））との間の距離の自乗であり、｛（POW^{DicS}（i））－（POW^{DicE}（i-1））｝^{2}となる。

【００２４】w_{POW}^{c}は、D_{POW}^{c}（u_i，u_{i-1}）に対する重み係数である。

【００２５】D_{FO}^{c}（u_i，u_{i-1}）は、i番目の音素片（u_i）の始端の基本周波数（FO^{DicS}（i））と、i－１番目の音素片（u_{i-1}）の終端の基本周波数（FO^{DicE}（i-1））との間の距離の自乗であり、｛（FO^{DicS}（i））－（FO^{DicE}（i-1））｝^{2}となる。

【００２６】w_{FO}^{c}は、D_{FO}^{c}（u_i，u_{i-1}）に対する重み係数である。

【００２７】D_{SPC}^{c}（u_i，u_{i-1}）は、i番目の音素片（u_i）の始端のスペクトル（SPC^{DicS}（i,j），j＝1～16）と、i－１番目の音素片（u_{i-1}）の終端のスペクトル（SPC^{DicE}（i-1,j），j＝1～16）との間の距離の自乗であり、｛（SPC^{DicS}（i,j））－（SPC^{DicE}（i-1,j））｝^{2}となる。

【００２８】w_{SPC}^{c}は、D_{SPC}^{c}（u_i，u_{i-1}）に対する重み係数である。

【００２９】入力された全ての音素に対するC_i^tとC_i^cとの総和C^{all}は、次の数式３で表される。

【００３０】

【数３】

$$C^{all} = \sum_{i}^{n} C_i^t + C_i^c$$

【００３１】

【発明が解決しようとする課題】ところで、上述したように音声合成方法によれば、品質の高い合成音声、つまり、自然発話に近い合成音声を得ることができる。しかしながら、自然発話から作成した音素片には、”なまけ”、”いいよどみ”など、実際に選択された場合に音質の劣化につながる音素片が存在している可能性が高い。このような音素片を含まないように波形辞書５を作成することが好ましいが、実際上には音質劣化につながる音素片をすべて取り除いて波形辞書５を作成することは困難である。

【００３２】また、波形辞書５を作成した後に、音質劣化につながる音素片を削除していくといったことも考えられるが、そのようにすると、波形辞書５の大幅な修正が必要となる。

【００３３】この発明は、波形辞書の大幅な修正を行なうことなく、音質劣化につながる品質の悪い音素片が最適な音素片として選択されにくくすることができる音声合成方法を提供することを目的とする。

【００３４】

【課題を解決するための手段】この発明による第１の音声合成方法は、複数の音声単位と各音素単位毎にターゲットとの歪みを算出するために用いられる補助情報とが波形辞書に格納されており、波形辞書に格納されている音素単位の組み合わせの中で、ターゲットとの歪みが最も少なくなる組み合わせを選択する音素単位選択型の音声合成方法において、各音素単位の補助情報にペナルティ情報を追加しておくステップ、ユーザが音声合成結果を聞いて、その品質が悪い場合には、品質の悪い合成音声箇所をユーザに入力させるステップ、ならびにユーザによって入力された品質の悪い合成音声箇所が入力された場合には、当該品質の悪い合成音声箇所に対応する音素片の補助情報内のペナルティ情報に、当該音素片が候補として選択されたときにターゲットとの歪み算出値を強制的に大きくさせるような値を設定するステップを備えていることを特徴とする。

【００３５】この発明による第２の音声合成方法は、複数の音声単位と各音素単位毎にターゲットに対する適応度を算出するために用いられる補助情報とが波形辞書に格納されており、波形辞書に格納されている音素単位の組み合わせの中で、ターゲットに対する適応度が最も大きくなる組み合わせを選択する音素単位選択型の音声合成方法において、各音素単位の補助情報に優先度情報を追加しておくステップ、ユーザが音声合成結果を聞いて、その品質が悪い場合には、品質の悪い合成音声箇所をユーザに入力させるステップ、ならびにユーザによって入力された品質の悪い合成音声箇所が入力された場合には、当該品質の悪い合成音声箇所に対応する音素片の補助情報内の優先度情報に、当該音素片が候補として選択されたときにターゲットに対する適応度の算出値を強

制的に小さくさせるような値を設定するステップを備えていることを特徴とする。

【００３６】

【発明の実施の形態】以下、この発明の実施の形態について説明する。

【００３７】〔１〕第１の実施の形態の説明

音声合成装置の全体構成は、図１と同じである。

【００３８】第１の実施の形態では、次の点（１）、（２）、（３）が従来と異なっている。

【００３９】（１）　図４に示すように、各音素片の補＊助情報に、ペナルティー情報$D'_{pri}(u_i)$を追加する。ペナルティー情報$D'_{pri}(u_i)$の初期値は、０である。

【００４０】（２）　音素単位選択部３で歪みC^{all}を算出するためのC_i'に、ペナルティー情報$D'_{pri}(u_i)$をパラメータとして加える。

【００４１】つまり、C_i'は、次の数式４で表わされる。

【００４２】

【数４】

$$C_i' = w'_{POW} D'_{POW}(t_i,u_i) + w'_{PO} D'_{PO}(t_i,u_i) + w'_{DUR} D'_{DUR}(t_i,u_i) + D'_{pri}(u_i)$$

【００４３】（３）　ユーザが音声合成結果を聞いて、その品質が悪い場合には、品質の悪い合成音声箇所を音声合成装置に入力するようにする。音声合成装置は、ユーザによって入力された品質の悪い合成音声箇所が入力された場合には、品質の悪い合成音声箇所に対応する音素片の補助情報内のペナルティー情報$D'_{pri}(u_i)$の値を、所定値αに設定する。

【００４４】この所定値αとしては、たとえば、数式１のC_i'の予想される最大値の約１００倍の値が用いられる。具体的には、任意数の文章を入力したときの数式１の最大値を実験により求めておき、その最大値の１００倍の値を、所定値αとして設定する。

【００４５】上記（１），（２），（３）のような変更を行なうことにより、ペナルティー情報$D'_{pri}(u_i)$の値としてαが設定されている品質の悪い音素片（u_i）が候補として選択された場合には、その音素片とターゲットとの歪みC_i'が、従来法に比べてα分だけ大きくなり、当該音素片（u_i）が最適な音素片として選択されにくくなる。

【００４６】上記実施の形態によれば、波形辞書内に品質の悪い音素片が存在している場合に、その音素片を削除するといった大幅な辞書の修正を行なうことなく、音素片の補助情報にペナルティー情報$D'_{pri}(u_i)$を追加するといった小規模な修正を行なうことによって、品質の悪い音素片を選択されにくくすることができるようになる。

【００４７】高品質の音声合成装置の場合には、波形辞書内の音素片格納部には６万個程度の音素片が格納されるため、音素片格納部の容量は数十ＭＢに及ぶが、波形辞書内の補助情報格納部の容量は数ＭＢというように、音素片格納部の容量の十分の１以下とである。このため、上記実施の形態のように補助情報格納部のみの修正を行なう方が容易である。また、音素片の削除に品質の改善を行なう従来方法では、波形辞書全てを置き換える必要があるが、上記実施の形態の方法では補助情報にペナルティー情報$D'_{pri}(u_i)$を追加するといった修※正のみであるため、波形辞書の一部の変更のみで修正が可能である。

【００４８】また、ユーザが自由に波形辞書から品質の悪い音素片を削除することにより、合成音声の品質を改善させることも考えられるが、音素の種類によってはその音素に対応する全ての音素片を削除してしまうおそれがある。そうすると、当該音素を含む文章に対して合成音声を生成できなくなる可能性がある。

【００４９】これに対して、上記実施の形態による方法では、たとえ、ある音素に対応する全ての音素片に対するペナルティー情報$D'_{pri}(u_i)$の値が所定値αに設定されたとしても、当該音素を音声合成する際には、その音素に対応する音素片の中で最適な音素片が選択されるため、当該音素に対して合成音声を生成することができるという利点がある。

【００５０】〔２〕第２の実施の形態の説明

第１の実施の形態においては、音素単位選択部３では、波形辞書に蓄積されている音素片の組み合わせの中で、歪みが少なくなる組み合わせを選択しているが、音素単位選択部として、波形辞書に蓄積されている音素片の組み合わせの中で、適応度が大きくなる組み合わせを選択するものが知られている。

【００５１】適応度S^{all}は、一般的に次の数式５で表される。

【００５２】

【数５】

$$S^{all} = \sum_i^n S_i' + S_i^c$$

【００５３】数式５においてS_i'は、ｉ番目の音素について辞書から抽出した音素片（u_i）と実際に使用する環境（ターゲットt_i）との間の類似度を示しており、次の数式６で表される。数式６中の各変数は、数式１中の変数と同じである。

【００５４】

【数６】

$$S_i' = w'_{POW}(1/D'_{POW}(t_i,u_i)) + w'_{PO}(1/D'_{PO}(t_i,u_i)) + w'_{DUR}(1/D'_{DUR}(t_i,u_i))$$

【００５５】また、数式５において、S_i'は、ｉ番目の音素について辞書から選択した音素片（u_i）の始端

と、ｉ－１番目の音素について辞書から選択した音素片（u_{i-1}）の終端との間の類似度を示しており、次の数式７で表される。数式７中の各変数は、数式２中の変数＊と同じである。

【００５６】

【数７】

$$S_i^c = w_{POW}^c(1/D_{POW}^c(u_i,u_{i-1}))+w_{F0}^c(1/D_{F0}^c(u_i,u_{i-1}))+w_{SPC}^c(1/D_{SPC}^c(u_i,u_{i-1}))$$

【００５７】第２の実施の形態では、次の点（１）、（２）、（３）が、適応度を用いて音素単位を選択する従来例と異なっている。

【００５８】（１）　各音素片の補助情報に、優先度情報E^t_{pri}（u_i）を追加する。優先度情報E^t_{pri}（u_i）の初期値は、所定値である。

【００６２】（３）　ユーザが音声合成結果を聞いて、その品質が悪い場合には、品質の悪い合成音声箇所を音声合成装置に入力するようにする。音声合成装置は、ユーザによって入力された品質の悪い合成音声箇所が入力された場合には、品質の悪い合成音声箇所に対応する音素片の補助情報内の優先度情報E^t_{pri}（u_i）の値を、初期値より小さい値に設定する。

【００６３】

【発明の効果】この発明によれば、波形辞書の大幅な修正を行なうことなく、音質劣化につながる品質の悪い音素片が最適な音素片として選択されにくくすることができる。

※【００５９】（２）　音素単位選択部３で適応度S^{all}を算出するためのS_i^tに、優先度情報E^t_{pri}（u_i）をパラメータとして加える。

【００６０】つまり、S_i^tは、次式８で表わされる。

【００６１】

【数８】

$$S_i^t = w_{POW}^t(1/D_{POW}^t(t_i,u_i))+w_{F0}^t(1/D_{F0}^t(t_i,u_i))+w_{DUR}^t(1/D_{DUR}^t(t_i,u_i))+E_{pri}^t(u_i)$$

【図面の簡単な説明】

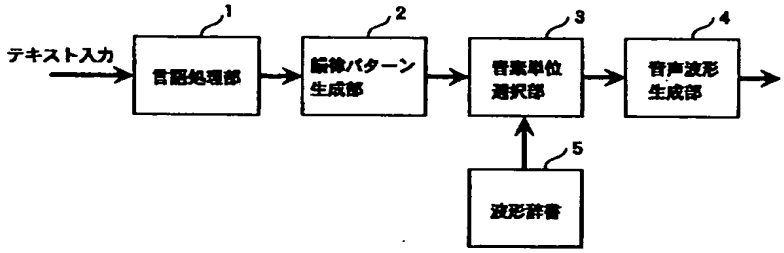
【図１】音声合成装置の全体構成を示すブロック図である。

【図２】波形辞書５の内容を示す模式図である。

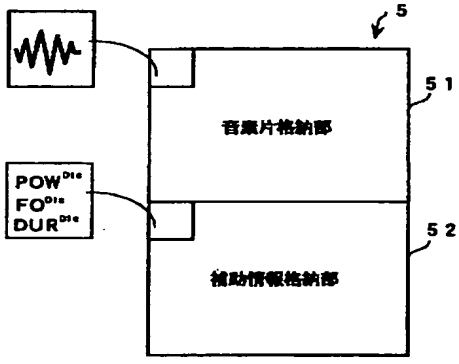
【図３】音素単位選択部３において、音素片の組み合わせを選択するために用いられる２種の歪みC_i^t、C_i^cを説明するための模式図である。

【図４】品質の悪い合成音声箇所に対応する音素片の補助情報に、ペナルティー情報D^t_{pri}を追加された様子を示す模式図である。

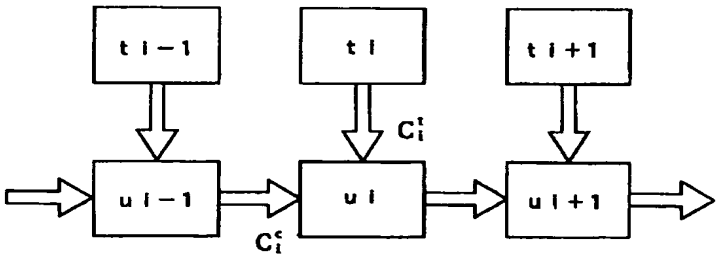
【図１】



【図２】



【図３】



【図４】